ЗН

УДК 681.3

А.О. Никоненко, К.С. Лиман, Є.В. Забєлін, Б.О. Рибачок
Київський національний університет ім. Т. Шевченка, Україна
Україна, 01601, м. Київ, вул. Володимирська, 64/13

UWN:Онтокоректор як інструмент створення лінгвістичної онтології української мови

A.O. Nikonenko, K.S. Lyman, E.V. Zabelin, B.O. Rybachok
Taras Shevchenko National University of Kyiv, Ukraine
Ukraine, 01601, Kyiv, Volodymyrska Street, 64/13

UWN:Ontocorrector as a Tool for Ukrainian Language Linguistic Ontology Creation

А.А. Никоненко, К.С. Лиман, Е.В. Забелин, Б.О. Рыбачок
Киевский национальный университет имени Тараса Шевченко, Украина
Украина, 01601, г. Киев, ул. Владимирская, 64/13

UWN:Онтокорректор как инструмент создания лингвистической онтологии украинского языка

Стаття присвячена питанню створення та практичного використання спеціалізованих утиліт у процесі формування онтологічних баз знань. Детально розглянуто один з підкласів онторедакторів – онтокоректори, показано їх місце в процесі створення онтології, основні задачі, особливості впровадження, практичні і методологічні рекомендації з використання.
Ключові слова: онтокоректор, онтологія української мови, створення онтологій.

In the article the task of authentication of parameter is examined as a continuous function of parabolic equalization is in partials. Analytical expression is found for the calculation of gradient of the non – obvious set functional. Gradient is used to determine the modernized classical method of Lagrange multipliers.
Key words: ontocorrector, Ukrainian language ontology, ontologies creation.

Статья посвящена вопросу создания и практического использования специализированных утилит в процессе формирования онтологических баз знаний. Подробно рассмотрен один из подклассов онторедакторов – онтокорректоры, показано их место в процессе создания онтологии, основные задачи, особенности внедрения, практические и методологические рекомендации по использованию.
Ключевые слова: онтокорректор, онтология украинского языка, создание онтологий.

Вступ

Одним з ключових моментів створення лінгвістичної онтології є етап розробки спеціалізованого програмного забезпечення для аналізу та редагування її даних. Найбільш поширеним типом програмного забезпечення для виконання такої роботи з обробки знань є онторедактори. Детальні огляди утиліт даного класу можна знайти у роботах [1-4]. У загальному випадку онторедактори дозволяють модифікувати більшу частину характеристик онтології, серед них: редагувати концепти, змінювати зв’язки між концептами, видозмінювати структуру онтології, обирати мову представлення знань та навіть створювати нові онтології.

Онтокоректори – це підклас онторедакторів, спеціалізований для аналізу та адаптації знань онтології. Створення знань та внесення змін в онтологічну базу вимагає використання утиліт цього класу. Під адаптацією ми будемо розуміти процес створення нової онтології на основі вже існуючої. Класичний приклад – створення онтологій для європейських мов у проекті EuroWordNet [5] та створення онтологій балканських мов у проекті BalkaNet [6] на основі лексико-семантичної англомовної бази знань WordNet [7]. Аналогічні дослідження з адаптації англомовних знань WordNet та їх внутрішньої структури до української мови ми проводимо в проекті UWN [8]. Архітектура, модель представлення знань, методи доступу до онтології, а також інші деталі створеної україномовної лексичної бази знань описано в роботі [9].

Перші спроби зі створення україномовної онтології нами проводилися ще у 2005 – 2006 роках, отриманий за час проекту досвід вказав на необхідність систематизації та автоматизації деяких елементів процесу. Як відповідь на ці потреби було розроблено нову модель представлення знань та створено ряд спеціалізованих утиліт для проведення роботи з наповненням онтології, ключову роль серед яких відіграє онтокоректор.

Робота над онтокоректором нами проводиться з 2010 року, загалом система пройшла шлях від версії 0.1 до 1.55 (актуальна на момент написання статті версія додатку). Протягом цього часу нові версії системи з'являлися на сайті проекту UWN [10] та проходили практичні випробування серед користувачів, що допомагають нам створювати онтологію. На даний момент робота над створенням онтології ще триває, проте практичний досвід вже сьогодні дозволяє стверджувати: в процесі адаптації онтології одне з ключових місць має займати створення онтокоректора. Далі в статті ми розглянемо деякі особливості створення та використання утиліти цього типу.

У проекті UWN онтокоректор виконує широкий перелік функцій, серед них:

– забезпечення користувачів інтерфейсом до знань онтології, підтримка процесів видачі, редагування та збереження інформації;

– підтримка циклічних процесів зміни стану знань після редагувань та перевірок;

– забезпечення послідовного редагування знань без виникнення колізій у великій кількості паралельних редакторських та модераторських сесій;

– перевірка коректності введених даних, у тому числі орфографічна перевірка;

– персоналізація оброблених даних – кожен набір змін має свого власника.

Використання онтокоректора дозволило проводити одночасну роботу з наповнення онтології великої кількості людей без додаткового розподілу синсетів між виконавцями, тобто редактор може обрати для роботи будь-який синсет, а система автоматично додасть до синсета дані про автора.

Подальше редагування синсету, аж до остаточного затвердження в якості коректного, може проводити лише його автор. Також функціонал онтокоректора використовується в розробленій нами схемі адаптації даних англомовної онтології, що передбачає багаторівневу модель уточнення даних, для перевірки якості створених редакторами знань.

На першому етапі дані синсету перекладаються автоматично; на другому етапі ці дані редагує людина з правами редактора; після закінчення роботи редактора, на третьому етапі, синсет перевіряється модератором; фінальна перевірка даних проводиться адміністратором.

Онтокоректор використовується на другому та третьому етапах роботи над синсетом. Затверджені адміністраторами синсети використовуються на етапі валідації онтології з метою виявлення порушень та неточностей у структурі її знань.

1 Принципи створення

«UWN:Коректор» є одним з прикладних додатків до системи UWN – проекту зі створення та покращення україномовної онтології на базі WordNet. Детальний аналіз вимог показав, що для побудови системи найбільш вдало можна використати дворівневу клієнт-серверну архітектуру. Основна мета системи – надати користувачам можливість роботи з онтологічними даними з БД UWN.

1.1 Структура серверної частини системи UWN

Основними структурними елементами для розміщення логіки та даних в СУБД Oracle виступають схеми. Кожна схема може містити таблиці з даними, модулі з програмною логікою, окремі процедури та функції, користувацькі типи даних, тригери та механізми запуску процесів за розкладом. Крім того, кожна схема може мати свої політики безпеки та доступу до об’єктів інших схем. Основні елементи програмної логіки та збереження даних UWN, а також зв’язки між ними показано на рис. 1. Більш детальний огляд серверної організації UWN можна знайти в статті [9].

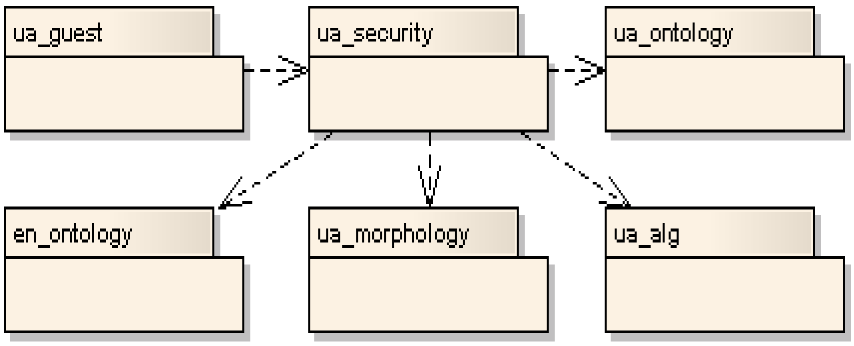


Рисунок 1 – Схема взаємодії основних структурних елементів UWN

Розглянемо призначення ключових блоків:

– ***ua_guest*** – технологічна схема, що використовується як єдина точка для підключення клієнтів і зовнішніх систем до БД UWN. Тобто, схема ***ua_guest***, що виконує роль буферної зони для системи безпеки;

– ***ua_security*** – схема, що відповідає за аутентифікацію систем, що підключаються до UWN, визначення набору повноважень та надання прав доступу відповідно до профілю системи-додатку. Також дана схема виконує роль єдиного інтеграційного вузла доступу до всіх внутрішніх інтерфейсів UWN (серверної логіки), що здійснює обробку та модифікацію даних онтологічної та морфологічної баз. Додатково, для підвищення рівня безпеки системи, схема містить механізми логування доступів (включаючи час, IP-адресу та інші характеристики машини-клієнта) та логування списку команд, що виконує додаток. Дії користувача також відслідковуються і запам’ятовуються іншими модулями БД, задля забезпечення контролю цілісності даних і для відновлення пошкоджених даних;

– ***ua_ontology*** – схема, що зберігає інформацію про наповнення україномовної онтології та містить серверну логіку для роботи з даними онтології (пошук та отримання даних синсету, зміна даних синсету);

– ***en_ontology*** – схема, що зберігає інформацію про наповнення англомовної онтології (аналогічна до ***ua_ontology)***;

– ***ua_alg*** – схема, що використовується для розміщення різноманітних семантичних алгоритмів та серверної логіки лінгвістичних додатків, а також програмно-алгоритмічних частин лінгвістичних наукових досліджень (наприклад, програмних реалізацій методів визначення міри семантичної зв'язності та схожості);

– ***ua_morphology*** – схема, що зберігає морфологічний словник української мови та методи перевірки правопису.

1.2 Механізм взаємодії клієнта і сервера

Захист системи орієнтовано на уникнення неавторизованого доступу до даних онтології. Механізм взаємодії клієнтського додатку з серверною частиною UWN містить наступні кроки:

1) для встановлення з'єднання з серверною частиною UWN клієнтський додаток входить до системи через технологічний аккаунт ***ua_guest*** і отримує право доступу до схеми ***ua_security*** для подальшої ідентифікації;

2) додаток проводить виклик процедур аутентифікації схеми ***ua_security*** та проходить реєстрацію в системі. Після успішної реєстрації клієнт отримує права на роботу з об'єктами UWN у відповідності до профілю додатку;

3) за умови успішно виконаного другого кроку система-додаток отримує права на виклик певної підмножини процедур та функцій з інших схем системи: ***ua_ontology, en_ontology, ua_alg, ua_morphology*** та інших;

4) на четвертому кроці відбувається безпосередня робота додатку з UWN, що може складатися як з виклику окремих лінгвістичних функцій, так і з взаємодії з пакетами серверної логіки (наприклад, таким чином побудовано системні утиліти).

1.3 Принципи організації роботи клієнтського додатку коректора

Види робіт у системі та права доступу користувачів регулюються спеціальними профілями, які зберігаються на сервері. При вході до системи користувач ідентифікується за допомогою логіну та паролю, після валідації сервером даних і аутентифікації в системі, клієнт отримує дані про рівень доступу, у відповідності з яким надається доступ до функціоналу клієнтського додатку. В залежності від профілю користувачі можуть мати право на:

– читання (тільки перегляд даних);

– редагування (можна вносити зміни до деяких полів);

– модерація (перевірка правильності, оцінка якості синсету, повернення на доопрацювання).

Кожен наступний рівень доступу включає в себе всі права та можливості попередніх рівнів. Для вирішення задач з покращення якості онтології система надає користувачу можливість перебирати синсети, слова яких містять помилки. При виборі такого синсету додаток виводить на користувацьку форму наступні дані:

– повний список слів, які входять до синсету;

– глосарій синсету (опис або смислове значення);

– номер синсету;

– поточна статистика користувача;

– інша інформація.

Виводяться як україномовний варіант синсету, доступний для редагування, так і його англомовний відповідник, що використовується у якості основи для перекладу. Слова синсету та слова з опису, що мають орфографічні помилки (відсутні у стандартно-

му словнику української мови), виділяються червоним кольором. Внесені зміни користувач може зберегти до БД, після чого автоматично отримує наступний синсет з помилками.

Важливим моментом при обробці синсету є встановлення міток коректності. Редактори можуть встановлювати мітки коректності для частин синсету (опис та слова), якщо вважають, що всі помилки в цих полях було виправлено. Модератори, окрім зміни даних міток, мають можливість встановлювати додаткові:

– «затверджено» – вказує на правильність перекладу синсету, проставляється у випадку, коли український варіант синсету повністю відповідає англійському оригіналу, синсет більше не надходить в обробку редакторам;

– «забраковано» – вказує на невідповідність між українським варіантом синсету та англійським оригіналом, необхідне доопрацювання редактором.

Схематичне зображення сценаріїв використання системи наведено на рис. 2.

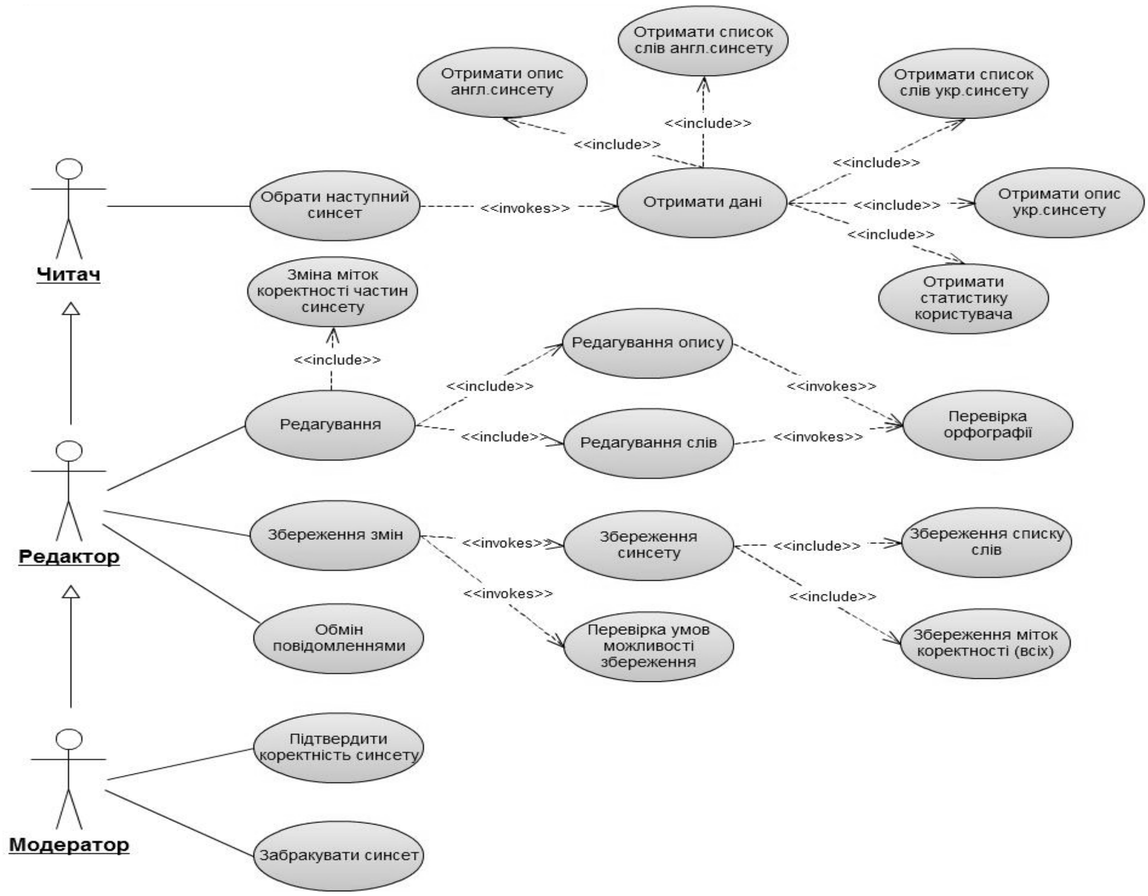


Рисунок 2 – Use-case діаграма UWN.
Зображено основні сценарії поведінки користувачів

1.4 Реалізація клієнтського додатку коректора

Клієнт відповідає за відокремлення логіки з БД та відображення інформації, тобто реалізує користувацький інтерфейс. Основні класи клієнтського додатку – це Manager, Synset та Word.

Manager відповідає за керування станом поточної сесії: *статистика і дані профілю користувача*, *дані про з'єднання з сервером*, *поточна частина мови*, *вказівник на поточний синсет*. Надає програмний інтерфейс до основних функцій сесії і запитів до БД:

– з'єднання з БД;

– авторизація і реєстрація користувача у системі;

– запит на отримання статистики користувача;

– перехід до наступного синсету (з БД надходить ідентифікатор поточного синсету, який передається в конструктор Synset, де вичитуються дані синсету з БД);

– збереження поточного синсету;

– запит на перевірку орфографії.

Synset інкапсулює дані, що описують наповнення синсету: *номер-ідентифікатор*, *частина мови*, *укр./англ. список слів*, *укр./англ. глосарії* та спеціальні поля: *мітки коректності* та *коментарі користувачів*. Містить методи управління даними полями, у тому числі методи додавання, вилучення та відновлення слів україномовного синсету (всі маніпуляції проводяться на стороні клієнта і зберігаються в БД лише при виконанні відповідного виклику). Також реалізовано методи для отримання даних синсету з БД та збереження змін на сервері.

Word містить дані про слово синсету: *ідентифікатор слова*, *мова*, *власне слово*, *мітки коректності* і спеціальне бінарне поле *isDeleted*, що має стани «вилучено» та «не вилучено» (зі списку слів синсету). Також має методи, що надсилають до БД запити на збереження і видалення слова.

Запити до БД в останніх двох класах як правило збігаються у часі зі створенням чи знищенням об'єкта класу. Тому всі проміжні зміни, внесені користувачем, і початкова конфігурація синсету, зберігаються на клієнтському рівні, що забезпечує можливості відміни редагування (функціональність додано у версії 1.4).

За відображення інформації і можливості редагування відповідає клас головної форми з деякими додатковими класами. Даний клас реалізує наступні функції:

– виділення помилок у полях, доступних для редагування, на основі результатів орографічної перевірки, здійсненної на стороні сервера;

– підтримка стеку дій користувача для відміни і повтору внесених змін;

– інформування користувача та реакція на виключні дії (exceptions) та помилки при роботі з БД;

– управління сесією (вхід, вихід, зміна користувача);

– правила доступу до функціонала в залежності від профілю користувача.

Користувацький інтерфейс додатку складається з двох основних форм: форма входу та робоча форма. Робоча форма UWN версії 1.55 представлена рис. 3.

У процесі розвитку та доопрацювання системи, за рекомендаціями користувачів, було додано наступні можливості:

– функціональність додавання і видалення слів з синсету, що пов'язано з відсутністю однозначного відображення між англійською та українською мовами;

– виділення орфографічних помилок;

– обмін повідомленнями між редактором і модератором, для проведення дискусій над помилками;

– інтерфейс визначення коректності слів було змінено на більш дружній для користувача;

– функціональність відновлення початкового стану синсету.

З системи було виключено надання орфографічних підказок через затримки при зверненні до сервера, де виконувався пошук схожих за написанням слів.

У наступних версіях клієнту плануються наступні зміни і доповнення:

– зміна навігаційної частини – замість послідовного перебору, список доступних синсетів (що особливо важливо для використання додатку в рамках профілю «Читач»);

– можливість переглядати контекстно-семантичну інформацію, щодо поточного синсету, його зв'язки з іншими синсетами, місце в таксономічній ієрархії, тощо;

– покращення системи обміну повідомленнями;

– покращення інформативності форми;

– оптимізація роботи з БД.

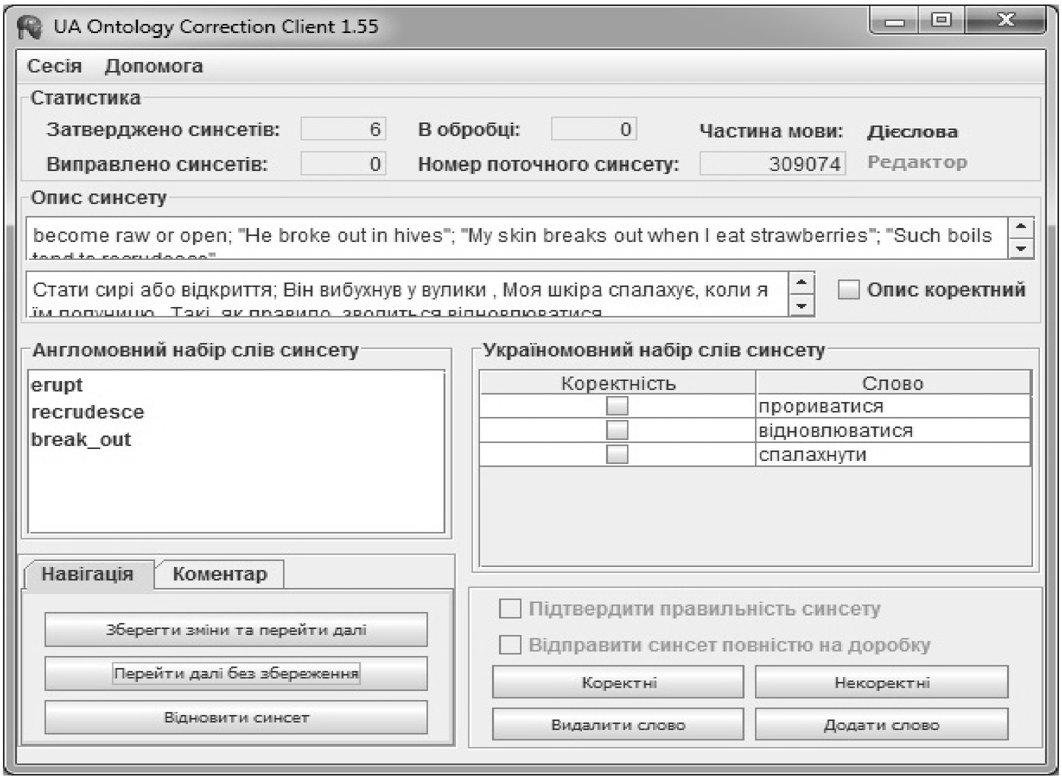


Рисунок 3 – Головна форма UWN:Коректора 1.55

2 Використання додатку в процесі адаптації онтології

Онтокоректор працює таким чином, що одразу після входу в систему редактор отримує синсет для роботи. Синсет, який було опрацьовано певним редактором, закріплюється за ним, подальші зміни в синсет може вносити тільки редактор-власник та модератор, що перевіряє даний синсет. Для опрацювання система видає користувачу в першу чергу ті синсети, що належать даному редактору і знаходяться в стані «некоректно», коли власні синсети редактора закінчуються, система видає користувачу нові синсети, що нікому не належать.

2.1 Редагування

При редагуванні синсету існує ряд чітких вимог, виконання яких забезпечує належну якість роботи:

1. Україномовний глосарій повинен описувати те ж саме поняття, що і англомовний. Глосарій повинен бути узгодженим за відмінкам, родами і числами. Також глосарій синсету може містити приклади використання поняття, що описується синсетом, наприклад:

Англійський глосарій: «a human being; "there was too much for one person to do"»

Український глосарій: «людина; "це було забагато, щоб виконати одній людині"»

2. Список слів україномовного синсету є аналогом списку слів англомовного синсету. Всі слова обов'язково повинні бути або синонімами, або близькими за змістом поняттями, що відповідають опису в глосарії. Крім цього, всі слова одного синсету мають належати до однієї частини мови. Всього в онтології присутні синсети чотирьох типів: іменники, дієслова, прикметники та прислівники. Зрозуміти, до якої частини мови належить синсет, можна за глосарієм. Наприклад, вище було наведено глосарії синсетів-іменників. Також в синсетах можуть зустрічатися слова, які перекладаються виключно у вигляді словосполучень. Наприклад:

Англійський глосарій: «a living organism characterized by voluntary movement»
Англійське слово: «animate_being»
Український глосарій: «живий організм, що характеризується свідомим рухом»
Українське слово: «жива_істота»

3. Допускається два варіанти написання словосполучень: через пропуск та через нижнє підкреслення, проте рекомендовано використовувати пропуск. Словосполучення потрібно використовувати тільки в тому випадку, якщо не існує можливості описати поняття одним словом. Наприклад:

plant_life = рослинне життя
natural_object = природний об'єкт
nutrient = поживна речовина

4. Окрема увага приділяється тому, щоб словосполучення не перетворювалися на речення, особливо на складні речення з різними типами зворотів. Наприклад:

особа, яка_укладає_документи = укладач документів
той, хто_проголошує_тост = проголошувач тосту
Марія,_королева_Шотландії = королева Шотландії Марія

Після закінчення роботи над синсетом редактор має зберегти зміни. Синсет вважається повністю обробленим, якщо глосарій і всі слова коректні. Такий синсет відсилається на перевірку модератору, який підтверджує правильність синсету, або повертає його назад на доопрацювання.

Всі внесені модератором правки відображаються редактору. Наприклад, якщо два з трьох слів синсету відмічено як некоректні, то синсет повертається автору-редактору з відповідними маркерами, що позначають некоректні слова, які потребують доопрацювання. Аналогічно працює маркер «опис коректний»: якщо його було знято модератором, то редактору буде показано відсутність маркера. В складних випадках модератор з редактором проводять дискусію у коментарях.

Закінчити роботу з синсетом можна двома методами: кнопкою «Зберегти зміни і перейти до наступного синсету», якщо внесені корективи необхідно зберегти в онтології; або кнопкою «Перейти до наступного синсету без збереження», якщо користувачу потрібно відмінити внесені зміни.

2.2 Модерація

Щодо модерації синсету існує ряд вимог та рекомендацій, основними серед яких є наступні:

1. При оцінці глосарію модератор повинен оцінити відповідність англійського і українського варіантів глосарія синсету, при цьому структура англійського глосарія повинна бути збережена в українському варіанті.

Варіанти роботи модератора з глосарієм:

– глосарій містить вірний за змістом переклад і являє собою стилістично коректне речення. В даному випадку немає необхідності вносити корективи;

– якщо глосарій містить незначні помилки (невірно узгодженні відмінки слів, пропущену літеру в слові і т.і.), то глосарій повинен бути виправленим модератором. Виключенням можуть бути лише глосарії, що містять надмірну кількість незначних помилок;

– невірний переклад або переклад з великою кількістю помилок. У даному випадку модератор знімає відмітку про коректність опису і синсет повертається назад редактору, що працював з ним.

2. Для оцінки слів модератору необхідно переглянути англомовний список слів та оцінити точність їх змістовного перекладу на українську мову. Кожному англійському слову повинно відповідати не менше одного україномовного варіанта.

Варіанти роботи модератора зі словами:

– якщо всі слова перекладено вірно, то необхідності вносити корективи немає;

– якщо у словах є незначні помилки, то модератору рекомендується самостійно виправити недоліки;

– у випадках, коли частину слів перекладено невірно, модератор має відмітити всі некоректні слова позначкою помилки;

– у випадках, коли всі слова перекладено невірно, модератор має скористатися спеціальною кнопкою «Коректних немає»;

– якщо і глосарій, і слова синсету перекладено невірно, то модератору рекомендується скористатися спеціальним маркером «відправити синсет повністю на доопрацювання». Після збереження глосарій і всі слова синсету буде позначено як некоректні, а сам синсет буде відправлено на повторну обробку автору-редактору.

Всі ухвалені модератором синсети вважаються вірними до того моменту, доки в них не знайдено помилки адміністратором системи.

3 Результати

Основним результатом проведеної роботи є онтологія української мови. На момент написання статті нами повністю опрацьовано всі синсети групи іменників – близько 80 тисяч одиниць. Наразі ведеться робота над синсетами інших частин мови: дієсловами, прикметниками та прислівниками.

В процесі роботи над онтологією виникло багато питань як методологічного, так і прикладного характеру. Методологічні питання, в основному, було зосереджено навколо принципів адаптації різних груп синсетів. До основних рішень даної категорії можна віднести наступні:

– збереження вихідної структури глосарія, в т.ч. збереження прикладів використання терміна і додаткової інформації про поняття, що описується синсетом;

– заборона використання складнопідрядних, складносурядних речень та інших складних структур у списку слів синсету. Допускається використання слів та словосполучень з двох і більше слів;

– кількість слів в україномовному варіанті синсету не може бути меншою за кількість слів в англомовному синсеті. З даного правила існують виключення, пов'язані з тим, що деякі поняття в англійській мові описуються більшою кількістю синонімів, ніж відповідні поняття в українській мові. Також існує ряд спеціальних ситуацій:

– прізвища російських або українських письменників, що мають декілька різних написань в англійській мові, і лише одне в українській;

– латинські терміни (назви тварин, рослин, елементи біологічної класифікації і т.д.) не підлягають перекладу;

– деякі слова англійської мови можна лише транслітерувати, реалії перекладаються на українську мову у вигляді опису поняття;

– трирівнева система контролю якості. Як показав наш досвід, якість отриманих знань більше залежить від правильності процесу оцінювання даних, ніж від рівня експертів, що корегують синсет. Справа в «людському факторі», адже навіть найбільш підготовлена людина допускає помилки. Тому необхідна багаторівнева система контролю, що має можливості для повернення синсетів на дообробку з наданням інформації про причини. Використання такої системи дозволяє значно знизити кількість помилок та виявити на ранньому етапі систематичні помилки користувача, що дозволяє запобігти їх виникненню в майбутньому;

– кількість експертів з контролю якості зменшується від рівня до рівня, а їх кваліфікація зростає, тому на верхньому рівні нами додатково використовуються автоматичні метрики та шаблони для виявлення помилок. Після перевірки експертом точності роботи такого шаблону на невеликій кількості даних, шаблон використовується на всьому масиві знань для автоматичного повернення на доопрацювання синсетів з помилками.

До питань прикладного характеру відносяться питання безпеки, способу доступу та особливості використання системи. Серед них:

– обов’язкова авторизація кожного користувача, як з метою забезпечення відповідності його профілю до рівня доступу, так і для можливості відстеження внесених змін. Дане рішення дозволяє автоматично виявити та відмінити одразу всі зміни, внесені несумлінними користувачем;

– захищений доступ до системи. В перших версіях клієнт-онтокоректор розроблявся у вигляді java-додатку, для забезпечення кросплатформеності рішення. Проте, java-додатки легко декомпілюються та розбираються на складові частини, не зважаючи на використання обфускації. Отримані таким чином функції для роботи з сервером можуть бути використані порушником для створення власних програм з нецільового використання онтології та для пошкодження даних;

– як випливає з попереднього пункту, система обов’язково повинна мати механізми відстеження та збереження змін, для повернення в попередній стан.

Висновки

Описана в статті утиліта дозволила в рамках проекту UWN реалізувати планомірний, контрольований та ефективний процес створення україномовної онтології. На даний момент виконано більше половини роботи з наповнення бази знань і здобуто важливий досвід, котрий дозволяє стверджувати, що протягом найближчих років роботу буде завершено. Дослідження онтології вже зараз вказують на необхідність проведення додаткового аналізу та, особливо, верифікації побудованої онтології з метою підвищення якості деяких її сегментів, що буде зроблено в наших наступних дослідженнях.

Література

1. Гладун А.Я. Онтологии в корпоративных системах. Часть II / А.Я. Гладун, Ю.В. Рогушина // «Корпоративные системы». – 2006. – № 1. – С. 48-56.
2. Gomez-Perez A. Ontology Engineering / A. Gomez-Perez, M. Fernando-Lopez, O. Corcho // Springer – Verlag. – 2004.
3. Handbook on Ontologies / [Staab S., Studer R. (eds).] // Springer – Verlag. – 2004.
4. Denny M. Ontology Tools Survey, Revisited / M. Denny [Електронний ресурс]. – 2004. – Режим доступу : http://www.xml.com/pub/a/2004/07/14/onto.html.
5. Piek Vossen(ed) EuroWordNet: A Multilingual Database with Lexical Semantic. – Networks Kluwer Academic Publishers, Dordrecht, 1998.

6. Grigoriadou, M., 2002. – Balkanet: A Multilingual Semantic Network for Balkan Languages / [Stamou S., Oflazer K., Pala K. ed all.]. : In Proceedings of the 1st Global WordNet Conference, Mysore, India.
7. George A. Miller. WordNet: A Lexical Database for English / George A. Miller // Communications of the ACM. – Vol. 38, № 11. – P. 39-41.
8. Никоненко А.О. Проект UWN: Досвід створення універсальної онлайн онтології української мови / А.О. Никоненко // Тези міжнародної наукової конференції ISDMCI'2011 «Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта». – Євпаторія, Крим, Україна, 2011. – С. 92-96.
9. Анісімов А.В. UWN: Універсальна онтологічна база знань української мови / О.О. Марченко, А.О. Никоненко // Проблеми програмування. – 2012. – № 2 – 3. – С. 348-355.
10. Сайт проекту UWN [Електронний ресурс]. – Режим доступу : http://lingvoworks.org.ua

Literatura

1. Gladun A.J.Ontologies for enterprise applications. Part II / A.J. Gladun, Y. Rogushina // magazine «Corporate Systems». – 2006. – № 1. – P. 48-56.
2. Gomez-Perez A., Fernando-Lopez M., Corcho O. Ontology Engineering // Springer –Verlag, 2004.
3. Staab S., Studer R. (eds). Handbook on Ontologies // Springer – Verlag, 2004.
4. Denny M. Ontology Tools Survey, Revisited // http://www.xml.com/pub/a/2004/07/14/onto.html. - 2004.
5. Piek Vossen(ed) EuroWordNet: A Multilingual Database with Lexical Semantic. Networks Kluwer Academic Publishers, Dordrecht, 1998.
6. Stamou S., Oflazer K., Pala K., Christodoulakis D., Cristea D., Tufis D., Koeva S., Tot-kov G., Dutoit D., Grigoriadou M. 2002. «Balkanet: A Multilingual Semantic Network for Balkan Languages». In Proceedings of the 1st Global WordNet Conference, Mysore, India.
7. George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, № 11: 39-41.
8. Nykonenko A. UWN Project: Ukrainian language universal online ontology creation experience // Proceedings of International Conference ISDMCI'2011 "Intelligent decision-making systems and the problems of computational intelligence", Evpatoria, Crimea, Ukraine, 2011 - P. 92-96.
9. Anisimov A.V., O.A. Marchenko, A. Nykonenko UWN: Universal ontological knowledge base of the Ukrainian language // Journal "Problems of Programming" – 2012. – № 2 – 3. – P. 348-355.
10. UWN Project webpage // http://lingvoworks.org.ua

RESUME

A.O. Nikonenko, K.S. Lyman, E.V. Zabelin, B.O. Rybachok

UWN:Ontocorrector as a Tool for Ukrainian Language Linguistic Ontology Creation

The article details the Ukrainian language ontology creating process using special ontocorrector tool. The experience gained in the ontology adaptation process, architecture basic principles and ideas are described. Mechanisms of interaction between client and server applications, program logic deployment, responsibilities and capabilities of each module are defined in detail. The structure of the system as a whole, its basic elements (up to the client application classes) and their purpose are considered. The main challenges and risks involved in making use of the utility are identified.

The ontology creation process using the ontocorrector described is the separate section. The main editing and validation principles are listed. Highly effective model of multi-level result analysis is offered and practically proven. Editing, moderating and administrating are key levels of this model. The article details the intent, responsibility and operation principles of each of the proposed levels. Frequent errors and adaptation specific complexities are further considered. The main value of the described system is its practical applicability, tested within the project UWN. This tool makes possible to implement a planned, controlled and efficient process of creating Ukrainian language ontology.

Стаття надійшла до редакції 22.04.2013.